Exhibit A

Pending Claims with Amendments Shown

24. (Amended) An array of polynucleotides, comprising:

a closely packed planar array of microparticles, the closely packed planar array having either a number of microparticles per unit area that is at least eighty percent of the number of microparticles in a hexagonal array of equal area or an average distance between centers of adjacent microparticles less than two microparticle diameters; and

a plurality of different polynucleotides attached to [~~each of~~] the microparticles such that each different polynucleotide is attached to a different microparticle.

25. (Cancelled) [~~The array of claim 24, wherein closely packed with reference to the planar array of microparticles requires either that the number of microparticles per unit area in the planar array is at least eighty percent of the number of microparticles in a hexagonal array of equal area or that the average distance between between centers of adjacent microparticles is less than two microparticle diameters.~~]

26. The array of claim 24, wherein the diameter of each of the microparticles is between about 0.1 μm and 100 μm.

27. The array of claim 24, wherein the plurality of different polynucleotides comprises a cDNA library.

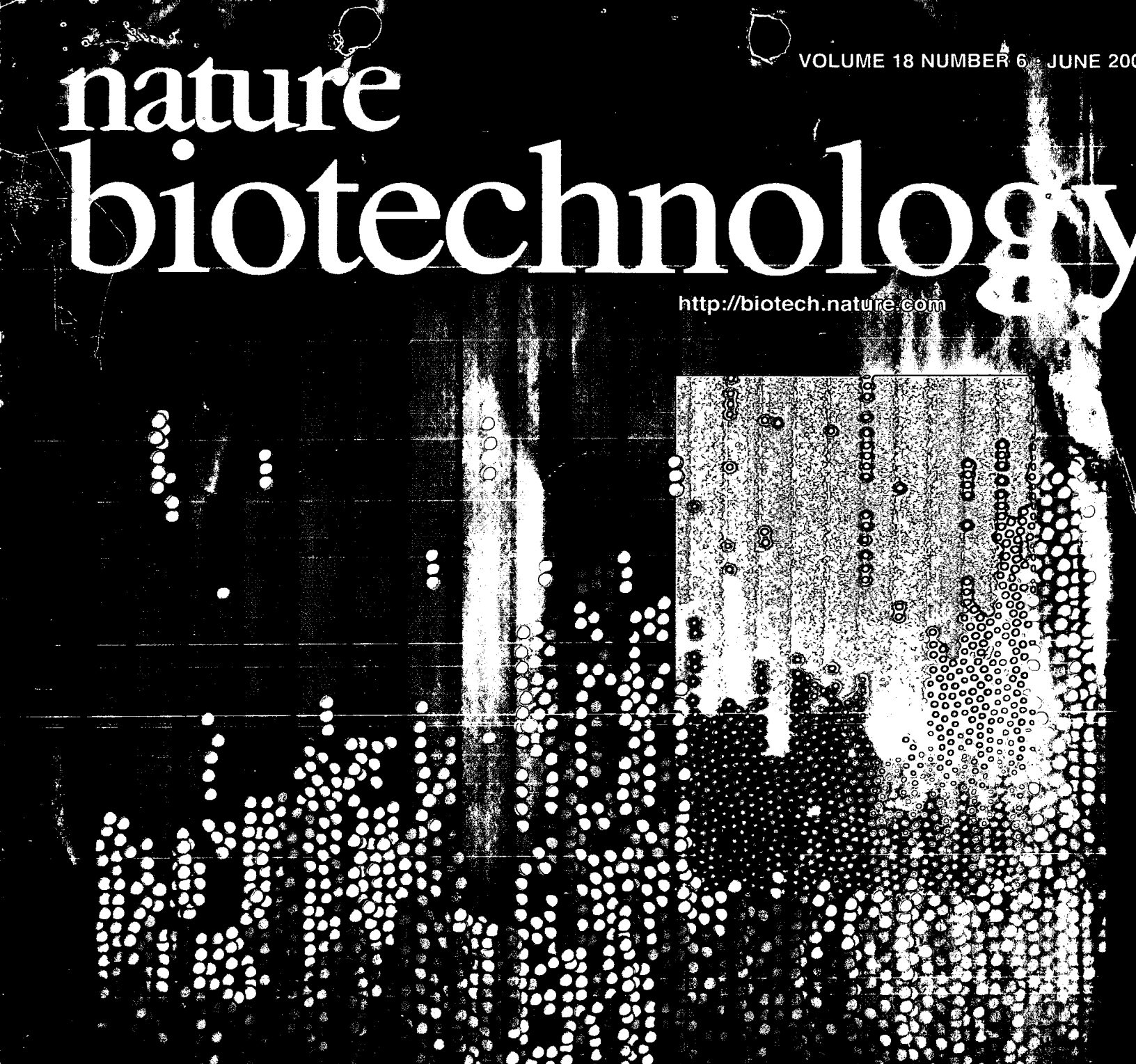
28. The array of claim 24, wherein the planar array of microparticles is disposed in a flow chamber.

--**35. (New)** The array of polynucleotides of claim 24, wherein said closely packed planar array has a number of microparticles per unit area that is at least eighty percent of the number of microparticles in a hexagonal array of equal area.--

VOLUME 18 NUMBER 6 · JUNE 200

nature biotechnology

http://biotech.nature.com



Massively parallel signature sequencing for expression analysis

Retroviral vectors for nondividing cells

Boosting carotene in tomatoes

Dopaminergic neurons from ES cells

Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays

Sydney Brenner*, Maria Johnson, John Bridgham, George Golda, David H. Lloyd, Davida Johnson, Shujun Luo, Sarah McCurdy, Michael Foy, Mark Ewan, Rithy Roth, Dave George, Sam Eletr, Glenn Albrecht, Eric Vermaas, Steven R. Williams, Keith Moon, Timothy Burcham, Michael Pallas, Robert B. DuBridge, James Kirchner, Karen Fearon, Jen-i Mao, and Kevin Corcoran

*Lynx Therapeutics, Inc., 25861 Industrial Blvd., Hayward, California 94545 *Corresponding author (e-mail: sbrenner@lynxgen.com).*

Received 17 February 2000; accepted 19 April 2000

We describe a novel sequencing approach that combines non-gel-based signature sequencing with in vitro cloning of millions of templates on separate 5 μm diameter microbeads. After constructing a microbead library of DNA templates by in vitro cloning, we assembled a planar array of a million template-containing microbeads in a flow cell at a density greater than 3×10^6 microbeads/cm^2. Sequences of the free ends of the cloned templates on each microbead were then simultaneously analyzed using a fluorescence-based signature sequencing method that does not require DNA fragment separation. Signature sequences of 16–20 bases were obtained by repeated cycles of enzymatic cleavage with a type IIs restriction endonuclease, adaptor ligation, and sequence interrogation by encoded hybridization probes. The approach was validated by sequencing over 269,000 signatures from two cDNA libraries constructed from a fully sequenced strain of *Saccharomyces cerevisiae*, and by measuring gene expression levels in the human cell line THP-1. The approach provides an unprecedented depth of analysis permitting application of powerful statistical techniques for discovery of functional relationships among genes, whether known or unknown beforehand, or whether expressed at high or very low levels.

Keywords: DNA sequencing, ligation, gene expression, fluid microarray, yeast

After the first complete sequence of a human genome is obtained, the next challenge will be to discover and understand the function and variation of genes and, ultimately, to understand how such qualities affect health and disease[1,2]. A key to this undertaking will be the availability of methods for efficient and accurate identification of genetic variation and expression patterns among large sets of genes[2]. Several powerful techniques have been developed for such analyses that depend either on specific hybridization of probes to microarrays[3,4] or on the counting of tags or signatures of DNA fragments[5–8]. Whereas the former provides the advantages of scale and the capability of detecting a wide range of gene expression levels, such measurements are subject to variability relating to probe hybridization differences and cross-reactivity, element-to-element differences within microarrays, and microarray-to-microarray differences[9-11]. On the other hand, the latter methods, which provide digital representations of abundance, are statistically more robust; they do not require repetition or standardization of counting experiments (since counting statistics are well modeled by the Poisson distribution), and the precision and accuracy of relative abundance measurements may be increased by increasing the size of the sample of tags or signatures counted[9]. Unfortunately, however, this property is difficult to realize routinely because of the cost and scale of effort required.

To address some of these problems, we describe a method for sequencing DNA that does not require physical separation of fragments and show how combining it with in vitro cloning of DNA templates on microbeads[12] results in a robust new analytical platform for genomic analysis. The power of this approach, which we refer to as massively parallel signature sequencing (MPSS) analysis, resides in the ability to conveniently handle complex mixtures of nucleic acid fragments by in vitro cloning of constituent fragments onto microbeads in sufficient quantities to conduct and monitor biochemical or enzymatic reactions by fluorescent probes. We show that multiple cycles of a ligation-based DNA sequencing method can be simultaneously carried out on a million microbeads, each having copies of a single template attached, to generate millions of signature sequences. Template-containing microbeads are assembled in a flow cell that constrains the microbeads to form a closely packed planar array that remains fixed as sequencing reagents are pumped through the flow cell. Sequencing progress is monitored optically by collecting and imaging fluorescent signals generated by the entire microbead array onto a CCD detector followed by image processing.

We show how MPSS analysis can be used to simultaneously acquire in a single operation hundreds of thousands of signature sequences from a yeast cDNA library, and we validate the accuracy of the signatures by comparison with the known genome sequence of *Saccharomyces cerevisiae*. We also demonstrate the technique's potential for gene expression analysis by comparing expression levels of genes of the human acute monocytic leukemia cell line, THP-1, measured by MPSS analysis and by conventional sequencing.

Results

In vitro cloning on microbeads. Before sequencing, templates are "cloned" on microbeads by first generating a complex mixture of conjugates between the templates and oligonucleotide tags, where the number of different oligonucleotide tags is at least a hundred times larger than the number of templates. For example, in the present

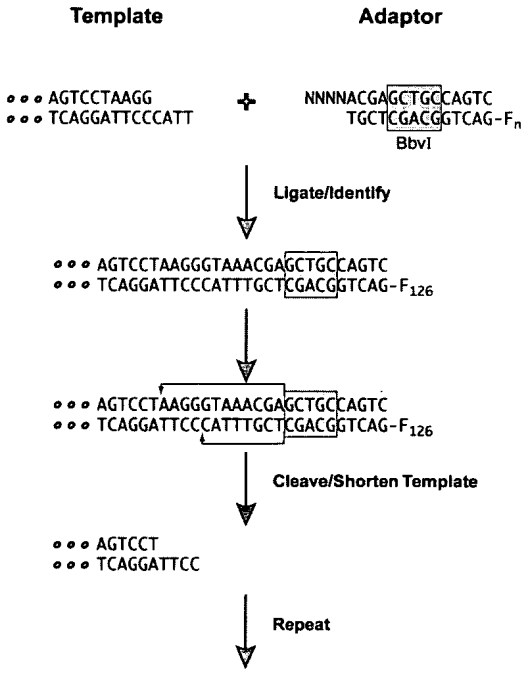


Figure 1. Ligation-based sequence determination using the type IIs restriction endonuclease *Bbv*I. A mixture of adaptors including every possible overhang is annealed to a target sequence so that only the one having a perfectly complementary overhang is ligated. Each of the 256 adaptors has a unique label, F_n, which may be detected after ligation. Above, the sequence of the template overhang is identified by adaptor label F_{126}, which indicates that the template overhang is "TTAC." The next cycle is initiated by cleaving with *Bbv*I to expose the next four bases of the template.

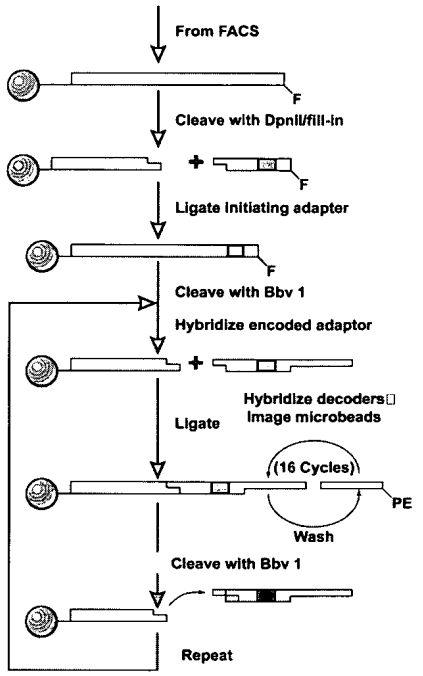


Figure 2. Use of encoded adaptors to identify four bases in each ligation–cleavage cycle. After microbeads loaded with fluorescently labeled (F) cDNAs are isolated by FACS, the cDNAs are cleaved with *Dpn*II to expose a four-base overhang, which is then converted to a three-base overhang by a fill-in reaction. Fluorescently labeled (F) initiating adaptors containing *Bbv*I recognition sites are ligated to the cDNAs in separate reactions, after which the microbeads are loaded into flow cells. cDNAs are then cleaved with *Bbv*I and encoded adaptors are hybridized and ligated. Sixteen phycoerythrin (PE)-labeled decoder probes are separately hybridized to the decoder binding sites of encoded adaptors and, after each hybridization, an image of the microbead array is taken for later analysis and identification of bases. The encoded adaptors are then treated with *Bbv*I, which cleaves inside the cDNA to expose four new bases for the next cycle of ligation and cleavage.

implementation, cDNA templates representing 3-4 × 10^4 different transcripts are inserted into a cloning vector containing a set of 1.67 × 10^7 different 32-mer oligonucleotide tags to form a set of 5-7 × 10^{11} conjugates. A sample of conjugates is taken that includes 1% of the total number of tags (~1.6 × 10^5 conjugates in the example), thereby ensuring that essentially every template in the sample is conjugated to a unique tag and that at least one of each of the 3–4 × 10^4 different cDNAs is represented in the sample with >99% probability. The sample is then amplified by PCR, after which the tags are rendered single stranded and combined under stringent hybridization conditions with a population of microbeads that have attached all the different complementary tags. Because the tags in the sample of conjugates make up only 1% of the total number of tags, only 1% of the microbeads will be "loaded" with template molecules. This 1% is concentrated into a library of loaded microbeads with a fluorescence-activated cell sorter (FACS). Each microbead in such a library has attached a population of about 10^4–10^5 identical copies of a single kind of template molecule.

Table 1. Sequences of encoded adaptors[a]

Common strand:
5′-GACTGGCAGCTCGT
Encoded adaptors for detecting base 1:
5′-**NNNA**ACGAGCTGCCAGTCcatttaggcg
5′-**NNNG**ACGAGCTGCCAGTCctgattaccg
5′-**NNNC**ACGAGCTGCCAGTCaccaatacgg
5′-**NNNT**ACGAGCTGCCAGTCcgctttgtag
Encoded adaptors for detecting base 2:
5′-**NNAN**ACGAGCTGCCAGTCggaacctgaa
5′-**NNGN**ACGAGCTGCCAGTCtgtgcgtgat
5′-**NNCN**ACGAGCTGCCAGTCaccgacattc
5′-**NNTN**ACGAGCTGCCAGTCattcctcctc
Encoded adaptors for detecting base 3:
5′-**NANN**ACGAGCTGCCAGTCcgaagaagtc
5′-**NGNN**ACGAGCTGCCAGTCtggtctctct
5′-**NCNN**ACGAGCTGCCAGTCtagcggactt
5′-**NTNN**ACGAGCTGCCAGTCggcgataact
Encoded adaptors for detecting base 4:
5′-**ANNN**ACGAGCTGCCAGTCgcatccatct
5′-**GNNN**ACGAGCTGCCAGTCcaactcgtca
5′-**CNNN**ACGAGCTGCCAGTCcacagcaaca
5′-**TNNN**ACGAGCTGCCAGTCgccagtgtta

[a]Four-base overhangs in bold and decoder binding sites in lowercase.

Principle of MPSS analysis. Template sequences are determined by detecting successful adaptor ligations (Fig. 1), and a signature is obtained by monitoring a series of such ligations on the surface of a microbead in a fixed position in a flow cell. The sequencing method takes advantage of a special property of a type IIs restriction endonuclease; namely, its cleavage site is separated from its recognition site by a characteristic number of nucleotides. Thus, a type IIs recognition site can be positioned in an adaptor so that after ligation, cleavage will occur inside the template to expose further bases for identification in the following cycle. In the present implementation, cDNA templates on microbeads were initially cleaved by *Dpn*II and the resulting ends converted to three-base overhangs, to be compatible with the initiating adaptors. Different initiating adaptors, whose type IIs restriction

Table 2. Accuracy of MPSS signatures for yeast.

Log phase	Clones sequenced	Signatures identified	Accuracy (%)
Early	126,678	115,685	91
Late	142,415	127,934	90
Totals	269,093	243,619	

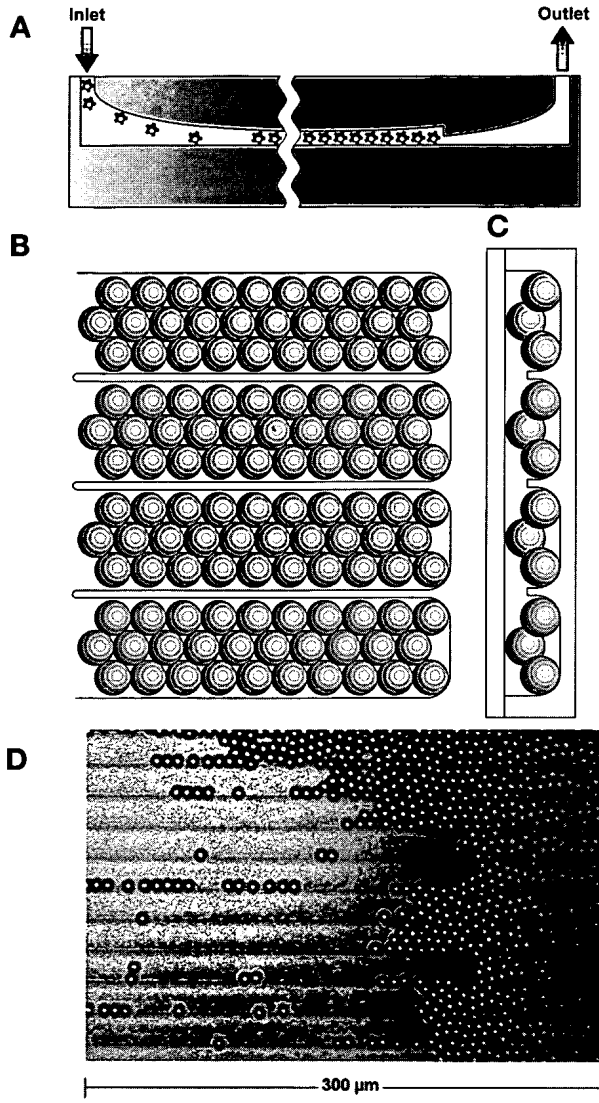


Figure 3. Flow cell design and use. The flow cell (A, longitudinal cross section) was fabricated by micromachining a glass plate to form a grooved chamber for immobilizing microbeads in a planar array (B, top view; C, lateral cross section). Microbeads in solution are loaded into the flow cell through the inlet, travel along the grooves (fluid flow from left to right in panel D), and finally pack against a vertical constriction, or dam, adjacent to the outlet to form a quasi-random array (D, right-hand side). Any minor displacements of microbeads that takes place after loading and during application of reagents are tracked by image processing software.

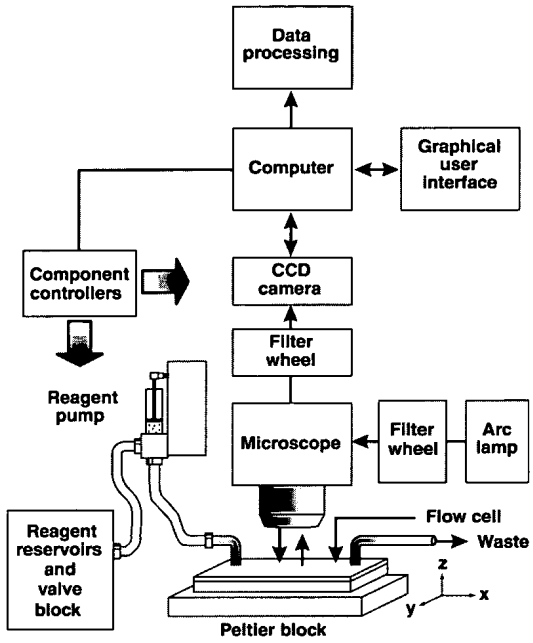


Figure 4. MPSS system. The flow cell was mounted on a confocal fluorescent microscope (Model BX60MF5; Olympus Optical, Tokyo) fitted with 75 W xenon arc lamp, motorized filter wheels, a computer-controlled stage (Ludl Electronic Products, Hawthorne, NY), a CCD camera Model PXL (Photometrics, Tucson, AZ) with a 2,000 × 2,000 pixel array (Kodak KAF4200), and a custom-made Peltier block integrated with the stage to control temperature. Reagent selection, flow rate, and flow cell temperature was controlled by a pentium-based computer programmed in LabVIEW (National Instruments, Austin, TX). For imaging, the microbead array was divided into 18 adjacent nonoverlapping subsections, each containing approximately 62,000 microbeads. For each subsection, two separate 10× images were collected: a fluorescent image and a reflected-light image for determining the center of each microbead during image processing. Image collection for each subsection took about 45 s and included the two imaging steps, two CCD data transfer steps, and a stage translation step. A 5 × 5 pixel array was assigned to each microbead image by image processing software such that the center of the microbead was contained in the central pixel of the array. A signal from a microbead was taken as the sum of values registered in the 9 central pixels of the 5 × 5 array in order to minimize contributions from adjacent microbeads. Raw data was collected for 20 bases from all microbeads in the flow cell in approximately five days. About 25% of the raw data sets yielded signatures after application of the base and signature calling algorithms. Data (not shown) from separate experiments demonstrate the per cycle efficiencies of adaptor ligation and cleavage within the flow cell to be about 80% and 90%, respectively, which are presently the primary contributors to overall loss of signatures in a sequencing run. Data (not shown) also indicate a significantly higher frequency of ambiguous base assignments among the higher numbered nucleotides of signatures, reflecting a fall in the signal-to-noise ratio in successive cycles.

sites were offset by two bases, were ligated to two sets of microbeads to reduce signature losses from self-ligation of ends of cDNAs produced when cleavage with *Bbv*I fortuitously exposes palindromic overhangs. Encoded adaptors (Fig. 2 and Table 1) were used that permit the identification of four bases in each cycle of ligation and cleavage, one base at a time. In each cycle, a full set of 1,024 encoded adaptors was ligated to the cDNAs, so that each microbead had an equal number of four different adaptors attached, one encoding the specific nucleotide at each position of the four-base overhang. The identity and ordering of nucleotides were then read off by specifically hybridizing, one at a time, each of 16 decoder probes to the decoder binding sites (lowercase bases in Table 1) of successfully ligated adaptors. The method continues with cycles of *Bbv*I cleavage, ligation of encoded adaptors, and decoder hybridization and fluorescence imaging.

To collect signature data, a microbead must be tracked through successive cycles of ligation, probing, and cleavage, a condition that is readily met by using a flow cell (Fig. 3) that constrains the microbeads to remain in a closely packed monolayer. Fluorescent signals from the microbead array were imaged onto a CCD (Fig. 4), where a digital representation of the microbead array was created. Image processing software was used to track positions of, and monitor fluorescent signals from, individual microbeads through successive hybridizations of decoder probes and through successive cycles of ligation and cleavage. False-color images of the microbead array (Fig. 5) display base calls in a color-coded format for any base position, and for each 20-base signature a collection of 65 separate fluorescent signals was collected for every microbead in the flow cell (see Experimental Protocol and Fig. 5 bar graph).

Signature accuracy was assessed by constructing cDNA libraries from mRNA extracted from early and late log phase yeast cultures, and subjecting them to MPSS analysis (Table 2). Of the 269,093 signatures called by the data processing algorithm, more than 90%

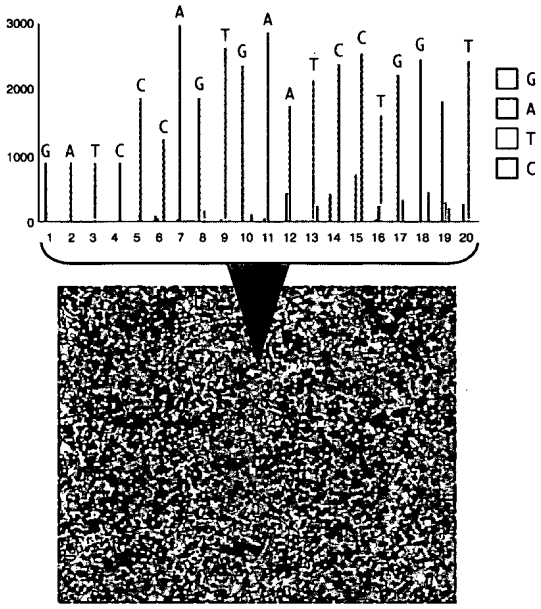


Figure 5. A false-color image of a portion of a microbead array with inset showing raw signature data from the microbead at the indicated position. The called base is shown above each histogram.

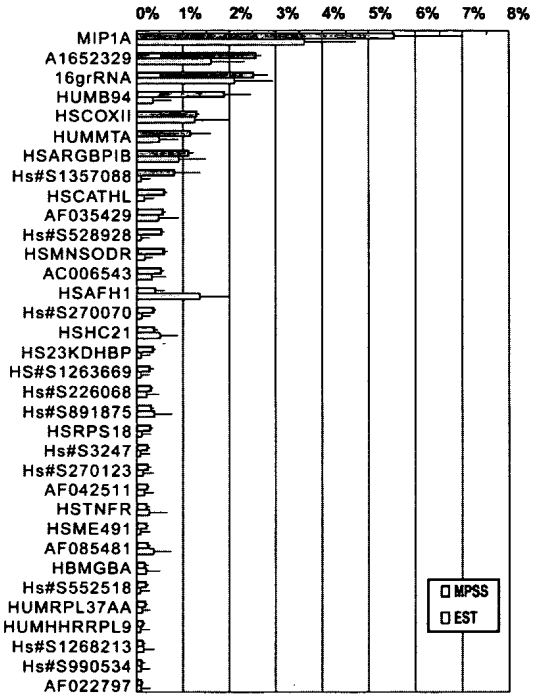


Figure. 6. Comparison of MPSS analysis with expressed sequence tag (EST) sequencing. A total of 1,839 randomly picked clones from a cDNA library derived from induced THP-1 cells were sequenced (PE Biosystems, Model 377 DNA Sequencer). The sequences were clustered and searched against GenBank with BLAST. The MPSS signatures corresponding to the EST clusters were then tabulated from the MPSS database of signatures from the same cDNA library. Percentage total for MPSS data is the average abundance measured, with actual measured error shown. Percentage total for the EST data is the number of sequences clustered out of the 1,839 selected clones, with the 99% confidence interval used for error. A 1% relative abundance corresponds to about 2,500 microbead signatures for MPSS data and to about 18 sequences for the EST data.

were identified in public yeast databases, which is comparable to a similar measurement by serial analysis of gene expression (SAGE) [13]. These results not only provide evidence of the accuracy of MPSS analysis, but also provide strong validation of the in vitro cloning technique. Without significantly pure populations of templates on the surfaces of the microbeads, few if any signatures would have been obtained. The majority of signatures without database matches were single copy, suggesting they were likely due to sequencing errors. Factors that may contribute to the spurious signatures include errors in the yeast genomic sequence, and errors introduced through reverse transcription, PCR, and incorrect ligation of encoded adaptors to noncomplementary overhangs or single-stranded tag complements on the microbeads.

The accuracy of gene expression measurements was also assessed by comparing expression levels of THP-1 genes measured by MPSS analysis and by conventional sequencing (Fig. 6). A database of over 1,619,000 signatures from MPSS analysis was generated from cDNAs derived from induced THP-1 cells. Separately, 1,839 clones were selected from the same cDNA library and conventionally sequenced. The relative frequencies of the most highly expressed genes were in substantial agreement and the measurement error from MPSS analysis was extremely low (Fig. 6), reflecting the advantage of large samples of templates. Reasons for the significant disagreement in a few of the expression measurements, such as apoferritin heavy-chain transcript (HSAFH1) and B94 protein mRNA (HUMB94), are being investigated.

Discussion

We have described a method for sequencing cDNAs cloned on the surfaces of microbeads that does not require physical separation of fragments to generate sequence information. Because the ligation-based method generates a time series of spatially localized signals, millions of microbeads carrying cDNAs can be assembled in closely packed microarrays for simultaneous analysis. Signatures of 16–20 bases are routinely generated, removing a source of ambiguity[14] associated with other digital methods of gene expression analysis. Although the methodology is still in its infancy, its bases-per-day throughput per machine is comparable to presently available high-capacity commercial sequencers, and its signatures-per-day throughput per machine exceeds that of such machines by over 10-fold. The defining performance characteristic of the MPSS approach is the generation of very large numbers of short read-length sequences: Whereas conventional sequencing machines process thousands of templates to give sequence read lengths of hundreds of bases, MPSS machines process millions of templates to give sequence read lengths of a few tens of bases. The parallelism of MPSS analysis is achieved at the molecular level; millions of templates are handled together in just a few reactions, without the need of separate isolation, processing of templates, or complex robotic systems. These characteristics make MPSS analysis particularly well suited for providing comprehensive assessments of gene expression.

Experimental protocol

Construction of oligonucleotide tag and anti-tag libraries, in vitro cloning, and formation of microbead libraries. Reagents and procedures used for in vitro cloning of cDNA templates on microbeads have been described elsewhere[12]. Briefly, a library of 32-mer anti-tags was synthesized by eight rounds of combinatorial addition of eight 4-mer subunits on glycidyl methacrylate microbead substrates (Bangs Laboratories, Fishers, IN). Approximately 10% of the anti-tags attached by a base-labile group were cleaved and used to construct a tag vector library into which cDNA derived from yeast or THP-1 cells was inserted to form tag–cDNA conjugate libraries. DNA was transformed into electro-competent *Escherichia coli* TOP10 cells (Invitrogen, Carlsbad, CA), which were grown in liquid cultures. For the microbead libraries, samples of 160,000 clones each were grown in 50 ml liquid cultures, after which tag–cDNA vectors were purified and tagged cDNAs were amplified using flanking PCR primers, one of which was fluorescently labeled. Tags of the amplified DNA were rendered single stranded as described[12], and 50 μg of the resulting mixture was combined with an aliquot of 16.7 million microbeads, each having about 10^6 copies of a single anti-tag, in a 100 μl reaction. The sample was incubated

for three days at 72°C, after which the microbeads were washed twice and the 1% microbeads having the brightest fluorescent signals were sorted on a MoFlo cytometer (Cytomation, Fort Collins, CO). Loaded, sorted microbeads were treated with T4 DNA polymerase in the presence of dNTP to fill in any gaps between the hybridized conjugate and the 5′ end of the anti-tag, after which the anti-tag was ligated to the cDNA by T4 DNA ligase.

Adaptors and decoder probes. Top strands of 16 sets of encoded adaptors (Table 1) were synthesized on an automated DNA synthesizer (PE Biosystems, Foster City, CA) where each set comprised 64 individual adaptors, and were separately combined with a common second strand to form double-stranded adaptors each having a single-stranded decoder binding site (lowercase) and a *Bbv*I recognition site positioned so that cleavage occurs immediately beyond the adaptor's four-base overhang. All 1,024 adaptors were combined in enzyme buffer (EB; 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.01% Tween 20). Then, 16 decoder probes were synthesized each having a sequence complementary to a different decoder binding site and a pyridyldisulfidyl R-phycoerythrin label (Molecular Probes, Eugene, OR) attached by a sulfosuccinimidyl 6-(3(2-pyridyldithio)propionamido)hexanoate crosslinker (Pierce, Rockford, IL) to an amino group (Clontech, Palo Alto, CA) attached through two polyethylene glycol (PEG) linkers to the 5′ end of the decoder oligonucleotide. Sixteen decoder probes were made (10 nM decoder in system buffer (SB), which consists of 50 mM NaCl, 3 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.9), 0.1% sodium azide). To initiate sequencing reactions by *Bbv*I cleavage at different positions along the cDNA templates offset by two bases, initiating adaptor 1 (5′-FAMssGACTGGCAGCTCGT, 5′-pATCACGAGCTGCCAGTC) and initiating adaptor 2 (5′-FAMssGACTGGCAGCAGTCGT, 5′-pATCACGACTGCTGCCAGTC) were synthesized, where "FAM" is 6-carboxyfluorescein (Molecular Probes), "s" is a PEG linker (Clontech), and "p" is phosphate (Clontech). To block ligation of encoded adaptors to free tag complements on the microbeads, cap adaptor (5′-DGGGAAAAAAAAAAAAAA, 5′-xTTTTTTTTTT) was synthesized, where x is a thymidylic residue (Glen Research, Sterling, VA) attached in reverse orientation to prevent concatenation of adaptors.

Sequencing DNA on microbeads. cDNAs on 2 million microbeads were digested with *Dpn*II (New England Biolabs, Beverly, MA) to provide a 5′-GATC overhang. After centrifugation and removal of the supernatant, the microbeads were treated with T4 DNA polymerase in the presence of 0.1 mM dGTP for 30 min at 12°C to create three-base overhangs on the free ends of the attached cDNAs. The microbeads were divided into two parts, and initiating adaptors 1 and 2 were separately ligated to different parts by combining 10^6 microbeads in 5 µl of TE (10 mM Tris, 1 mM EDTA) and 0.01% Tween 20 with 3 µl 10× ligase buffer (New England Biolabs), 5 µl adaptor in EB (25 nM), 2.5 µl T4 DNA ligase (2000 U μl^{-1}), and 14.5 µl distilled water, and incubating at 16°C for 30 min, after which the microbeads were washed three times in TE (pH 8.0) with 0.01% Tween. After resuspension in TE with 0.01% Tween, 10^6 microbeads of each part were loaded into separate flow cells where they were processed identically.

Reagents were pumped through the flow cells at a rate of 1 µl min^{-1}. SB was applied for 15 min at 37°C and for 15 min at 25°C, after which cap adaptor (1 nmol μl^{-1} in EB, T4 DNA ligase (Promega, Madison, WI) at 0.75 U μl^{-1}) was twice applied for 25 min at 16°C, first followed by SB for 10 min, Pronase wash (0.14 mg ml^{-1} Pronase (Boehringer, Indianapolis, IN) in PBS (Life Technologies, Rockville, MD) with 1 mM $CaCl_2$) for 25 min, and SB for 20 min, all at 37°C; and second followed by SB for 10 min, Pronase wash for 25 min, salt wash (SB with 150 mM NaCl) for 10 min, and SB for 10 min, all at 37°C. The microbeads were then imaged and positions in the flow cells recorded, after which three cycles of the following steps were carried out: *Bbv*I (1 U μl^{-1} in EB with 1 nmol μl^{-1} of carrier DNA: 5′-AGTGAACCTCGTTAGCCAGCAATC) was applied for 30 min, followed by SB for 10 min, Pronase wash for 25 min, salt wash for 10 min, and SB for 10 min, all at 37°C. Ligation mix (1 nmol μl^{-1} encoded adaptor, 0.75 U μl^{-1} T4 DNA ligase in EB) was twice applied for 25 min at 16°C, first followed by SB for 10 min, Pronase wash for 25 min, and SB for 20 min, and second followed by SB for 10 min, Pronase wash for 25 min, and SB for 10 min, all at 37°C. Kinase mix (0.75 U μl^{-1} T4 DNA ligase, 7.5 U μl^{-1} T4 polynucleotide kinase (New England Biolabs) in EB) was applied for 30 min at 37°C, followed by SB for 10 min, Pronase wash for 25 min, salt wash for 10 min, and SB for 10 min, all at 37°C. SB was applied for 75 min at temperatures varying between 20°C and 65°C, after which each decoder probe was successively applied for 15 min at 20°C, each application being followed by SB for 10 min at 20°C, microbead imaging with flow stopped, 100 mM dithiothreitol in SB for 10 min, and SB alone for 10 min both at 37°C. Each cycle was completed by applying SB for 10 min, Pronase wash for 25 min, salt wash for 10 min, all at 37°C, followed by SB for 10 min at 55°C and for 15 min at 20°C.

Base and signature calling. Raw data for a signature consists of a set of 65 fluorescence intensity values: 64 that consist of the 16 groups of four measurements from the interrogation of each base position by decoder probes for A, C, G, and T, over the four cycles, and a single fluorescence measurement of a signal generated by the initiating adaptor that is assigned to each nucleotide in the initial GATC overhang. After subtracting noise, a base (A, C, G, or T) was assigned to a position if its value was at least three times the next highest value and was above a predetermined minimum value. If the latter condition was met for the highest and next highest values, but not the former, then a two-base ambiguity code (R, Y, M, K, S, or W) was called if both values were above the predetermined minimum. Only a single ambiguous base was allowed per signature. Signatures were searched for homology in three yeast databases using the National Center for Biotechnology Information (NCBI) BLASTN Version 2.0[15] with default parameters, unless an ambiguous base was present in the signature. In the latter case, BLASTN was used with the word size parameter set to 7. The SGD open reading frame DNA database[16] was searched first, and a match was recorded if at least 16 consecutive bases matched those of a database sequence. If no matches were found for a signature, the NCBI yeast genomic database was then searched, and if still no matches were recorded, the NCBI nonredundant nucleotide database, nt, was searched.

Cell culture. *Saccharomyces cerevisiae* strain S288C (ATTC No. 204508) was grown as described[17]. Briefly, strain S288C was grown with orbital shaking at 30°C in YPD medium. Early and late log phase cultures were harvested at densities of $A_{600} = 0.6$ and $A_{600} = 3.2$, respectively. Cells were disrupted by repeated vortexing in the presence of lysis buffer (Novagen, Madison, WI) containing 500 µm glass beads (Sigma, St. Louis, MO), after which mRNA was purified from the lysate using a Straight A's mRNA isolation system (Novagen). THP-1 cells (ATCC No. TIB-202) were grown in DMEM/F12 medium supplemented with 10% heat-inactivated fetal bovine serum and induced by phorbol myristate acetate and lipopolysaccharide treatment as described elsewhere[12].

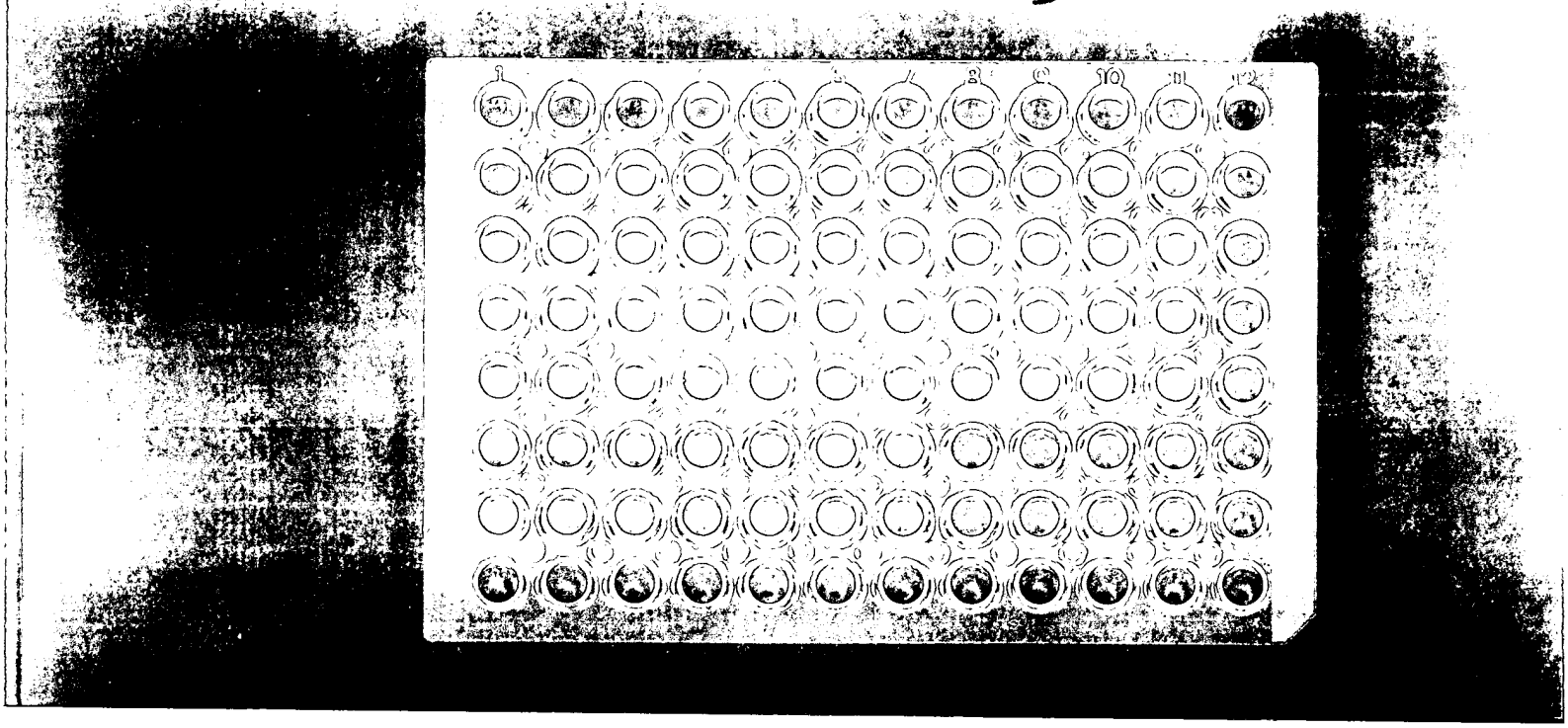
Acknowledgments

The authors thank Steve Macevicz of Lynx Therapeutics for preparing the manuscript; Mel Kronick of Agilent Technologies and Dan Pinkel of the Cancer Center, Department of Laboratory Medicine, University of California, San Francisco, for helpful comments; and Larry DeDionisio and Victor Quijano of Lynx Therapeutics for technical assistance.

1. Lander, E.S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).
2. Collins, F.S. et al. New goals for the U.S. human genome project: 1998–2003. *Science* **282**, 682–689 (1998).
3. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J.M. Expression profiling using cDNA microarrays. *Nat. Genet.* **21**, 10–14 (1999).
4. Hacia, J.G. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat. Genet.* **21**, 42–47 (1999).
5. Okubo, K. et al. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* **2**, 173–179 (1992).
6. Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
7. Bachem, C.W.B. et al. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J.* **9**, 745–753 (1996).
8. Shimkets, R.A. et al. Gene expression analysis by transcript profiling coupled to gene database query. *Nat. Biotechnol.* **17**, 798–803 (1999).
9. Audic, S. & Claverie, J. The significance of digital gene expression profiles. *Genome Res.* **7**, 986–995 (1997).
10. Wittes, J. & Friedman, H.P. Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J. Natl. Cancer Inst.* **91**, 400–401 (1999).
11. Richmond, C.S., Glasner, J.D., Mau, R., Jin, H. & Blattner, F.R. Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27**, 3821–3835 (1999).
12. Brenner, S. et al. *In vitro* cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA* **97**, 1655–1670 (2000).
13. Velculescu, V.E. et al. Characterization of the yeast transcriptome. *Cell* **88**, 243–251 (1997).
14. Neilson, L. et al. Molecular phenotype of the human oocyte by PCR–SAGE. *Genomics* **63**, 13–24 (2000).
15. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
16. Chervitz, S.A. et al. Using the *Saccharomyces* Genome Database (SGD) for analysis of protein similarities and structure. *Nucleic Acids Res.* **27**, 74–78 (1999).
17. Brewster, N.K., Val, D.L., Walker, M.E. & Wallace, J.C. Regulation of pyruvate carboxylase isozyme (*PYC1, PYC2*) gene expression in *Saccharomyces cerevisiae* during fermentative and nonfermentative growth. *Arch. Biochem. Biophys.* **311**, 62–71 (1994).

Exhibit C

Conventional 96-well plate (actual size)



Flow chamber containing an array of polynucleotides on microparticles (actual size)

Ligation of Linkers or Adapters to Double-Stranded cDNA

Linkers or adapters can be ligated to double-stranded cDNA (*UNIT 5.5*) to provide restriction endonuclease sites used in the production of a cDNA library (*UNIT 5.8*). For cloning purposes, only one linker or adapter must be present on each end of the cDNA. However, multiple linkers are usually ligated to the cDNA because both ends are phosphorylated and contain cohesive sequences. As a result, cDNA must first be methylated to protect it from a subsequent restriction digest designed to remove the multiple linkers (basic protocol). The procedure for ligating adapters to the cDNA is much simpler than that for linkers because only one end is phosphorylated, resulting in the ligation of just one adapter (alternate protocol). Linkered or adapted cDNA is then passed over a Sepharose CL-4B column to remove unligated linkers or adapters and other low-molecular-weight material (<350 bp) that would interfere with cloning (support protocol). This double-stranded cDNA is concentrated by ethanol precipitation and may be cloned directly (*UNIT 5.8*) or further fractionated by agarose gel electrophoresis (*UNITS 2.5 & 2.6*).

METHYLATION OF cDNA AND LIGATION OF LINKERS

BASIC PROTOCOL

To convert blunt-ended, double-stranded cDNA into DNA suitable for ligation to a vector, it is methylated by *Eco*RI methylase, ligated to *Eco*RI linkers, and digested with *Eco*RI. The methylation step protects *Eco*RI sites in the cDNA from *Eco*RI digestion. Linkered cDNA is then purified as described above.

Materials

Blunt-ended, double-stranded radiolabeled cDNA (*UNIT 5.5*)
2× methylase buffer (Table 3.1.3)
50× *S*-adenosylmethionine (SAM), freshly prepared (or from New England Biolabs with order of methylase)
*Eco*RI methylase (New England Biolabs; *UNIT 3.1*)
TE buffer (*APPENDIX 2*)
Buffered phenol (*UNIT 2.1*)
Diethyl ether
7.5 M ammonium acetate
95% and 70% ethanol, ice-cold
10× T4 DNA ligase buffer (*UNIT 3.4*) containing 5 mM ATP
1 μg/μl phosphorylated *Eco*RI linkers, 8- or 10-mers (Collaborative Research)
T4 DNA ligase (measured in cohesive-end units; New England Biolabs; *UNIT 3.14*)
*Eco*RI restriction endonuclease and 10× buffer (*UNIT 3.1*)
10× loading buffer without Xylene Cyanol (*UNIT 2.5A*)
CL-4B column buffer
Agarose, electrophoresis-grade
TBE electrophoresis buffer (*APPENDIX 3*)
Ethidium bromide solution (*UNIT 2.5A*)
DNA molecular weight markers (*UNIT 2.5A*)
10 mg/ml tRNA
3 M sodium acetate
65°C water bath
5-ml CL-4B column (see support protocol)

Additional reagents and equipment for cDNA synthesis (*UNIT 5.5*), quantitation of DNA (*APPENDIX 3*), agarose gel electrophoresis (*UNIT 2.5A*), and fragment purification (*UNIT 2.6*)

NOTE: See discussion of preparation of buffer stocks in reagents and solutions section.

Methylate blunt-ended, double-stranded cDNA

1. Dissolve blunt-ended, double-stranded cDNA pellet from *UNIT 5.5* in 23 µl water and add in the following order (50 µl final volume):

   25 µl 2× methylase buffer (1× final)
   1 µl 50× SAM (20 µg/ml final).

   Mix gently by pipetting up and down with a pipettor. Add 1 µl (20 U) *Eco*RI methylase to 400 U/ml final, mix as above, and incubate 2 hr at 37°C.

2. Add 150 µl TE buffer and extract with 200 µl buffered phenol beginning with the vortexing as in step 5 of the cDNA synthesis protocol in *UNIT 5.5*. Back extract the phenol phase with 100 µl TE buffer and pool the aqueous phases as described in step 6 of the cDNA synthesis protocol in *UNIT 5.5*.

   The cDNA, if prepared as described in the cDNA synthesis protocol, may be followed at all stages with a hand-held radiation monitor.

3. Extract 300 µl of aqueous phase twice with 1 ml diethyl ether as described in step 7 of the cDNA synthesis protocol in *UNIT 5.5*.

4. Ethanol precipitate with 125 µl of 7.5 M ammonium acetate and 950 µl of 95% ethanol, then wash with ice-cold 70% ethanol as described in steps 8 and 9 of the cDNA synthesis protocol in *UNIT 5.5*.

Ligate EcoRI linkers

5. Dissolve DNA in 23 µl water and add in the following order (30 µl final volume):

   3 µl 10× T4 DNA ligase buffer containing 5 mM ATP (1× and 0.5 mM ATP final)
   2 µl 1 µg/µl phosphorylated *Eco*RI linkers (67 µg/ml final).

   Mix gently by pipetting up and down with a pipettor. Add 2 µl (800 U) T4 DNA ligase to 27,000 U/ml final, mix as above, and incubate overnight at 4°C.

6. Microcentrifuge the ligation briefly and place tube in a 65°C water bath 10 min to inactivate the ligase.

Digest with EcoRI

7. Place tube on ice 2 min, then add in the following order:

   95 µl H_2O
   15 µl 10× *Eco*RI buffer (1× final).

   Mix gently by pipetting up and down with a pipettor. Add 10 µl (200 U) *Eco*RI restriction endonuclease to 1300 U/ml final, mix as above, and incubate 4 hr at 37°C.

   During incubation, prepare the CL-4B column described in the support protocol.

8. Add an additional 3 µl (60 U) *Eco*RI restriction endonuclease to the cDNA, mix, and incubate another hour at 37°C to ensure complete digestion of the linkers.

9. Place the tube containing the reaction mixture in a 65°C water bath 10 min to inactivate the endonuclease.

Remove excess linkers

10. Add 2 µl of 10× loading buffer without Xylene Cyanol to the reaction and load the cDNA onto a 5-ml CL-4B column prepared in a 5-ml disposable plastic pipet.

11. Allow the loaded sample to enter the column just until the top of the gel becomes dry. Fill the pipet with CL-4B column buffer and allow the column to flow by gravity, collecting ~200-µl fractions manually. Follow the cDNA with a hand-held radiation monitor; the bromphenol blue indicates the position of digested linkers. Stop collecting fractions after the main peak of counts has eluted and before the dye begins to elute.

12. Count 2-µl aliquots of each fraction and plot the results—the elution profile should appear similar to what is shown in Figure 5.6.1.

13. Pool the first one-half of the peak (save the rest as an ethanol precipitate, just in case); add 2.5 vol ethanol (using two or three microcentrifuge tubes as necessary), mix, and place 15 min on dry ice.

    The column buffer has sufficient NaCl for precipitation and no more should be added. If the cDNA is not to be ligated or further size-fractionated immediately, store it as an ethanol precipitate.

14. Remove the tubes from dry ice, let thaw, and microcentrifuge 10 min at full speed, 4°C; remove most of the supernatant, fill the tubes with ice-cold 70% ethanol, and microcentrifuge again. Remove most of the supernatant and dry the pellets under vacuum. Resuspend pellets in a total of 50 µl TE buffer.

15. Determine the cDNA concentration. Determine the amount of ^{32}P in 1 µl by scintillation counting and fractionate 2.5 µl on a 1% agarose minigel (see *UNIT 2.5*) to check the average cDNA size by ethidium fluorescence or autoradiography. A significant fraction of the double-stranded cDNA should be larger than 1.5 kb. If the double-stranded cDNA is to be cloned with no further size fractionation, proceed to ligation protocols in *UNIT 5.8*; otherwise, continue with step 16.

    Approximately 50% to 70% of the starting radioactivity (1 to 3 µg of cDNA) should be recovered and most of the cDNA should be >1.5 kb.

    Since only 50 to 100 ng of cDNA are required to produce a full complexity library with a phage vector, it is recommended that a library be produced at this stage in any event.

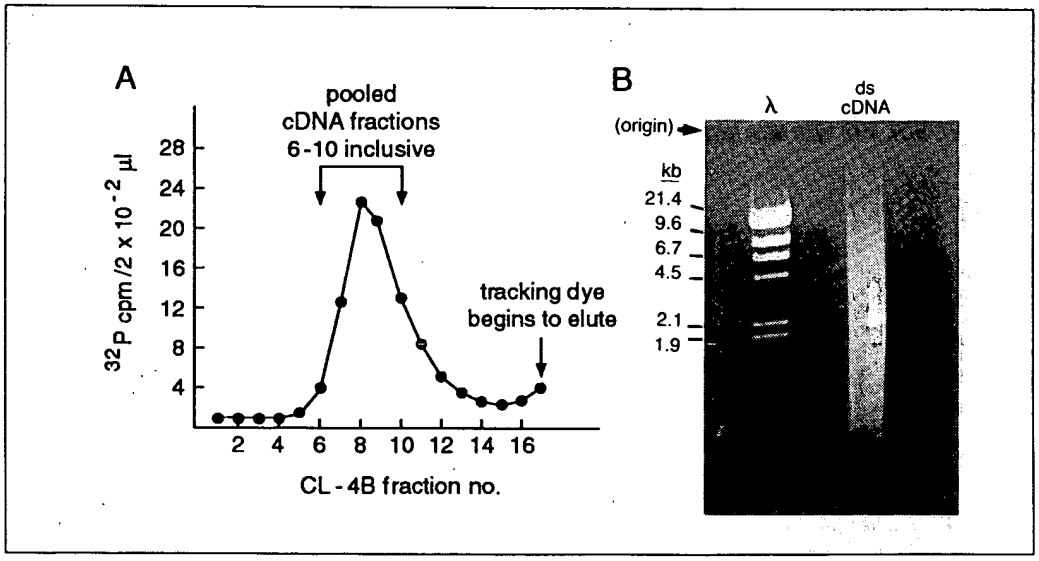


Figure 5.6.1 Fractionation of *Eco*RI-digested cDNA by **(A)** Sepharose CL-4B chromatography and **(B)** agarose gel electrophoresis.

A library may be stored for years and may be useful in the future. For example, sequences related to the gene of interest may be identified that were excluded from a size-fractionated library.

Size-select the cDNA to obtain long inserts

16. Pour a 0.8% TBE agarose minigel—the gel should be thick enough such that all of the cDNA will fit into a single well. Rinse the gel box, tray, and comb thoroughly, and use fresh TBE electrophoresis buffer.

    High-quality, nuclease-free agarose that does not inhibit ligation is essential. Most commercial agarose advertised as molecular biology grade is adequate. The agarose may be checked by first carrying an EcoRI fragment of a plasmid through the procedure and comparing its cloning efficiency in the cDNA vector, expressed as recombinants/ng insert, to the cloning efficiency of the same fragment prior to fractionation. Wash the gel box thoroughly afterward!

17. Add 10× loading buffer to 1× final to the cDNA and load it into a well near the center. Load DNA molecular weight standards (e.g., an *Hin*dIII digest of λ phage) two wells away from the sample. Electrophorese at 70 V until adequate resolution is achieved as determined by ethidium bromide fluorescence, usually 1 to 4 hr.

    Be sure not to use standards with EcoRI ends.

18. Elute double-stranded cDNA of the desired size as estimated by comparison with the comigrated standards.

    λgt10 and λgt11 have a maximum insert size of 7 kb, so collecting cDNA larger than this won't be useful unless a plasmid vector or a phage vector such as Charon 4A or EMBL 4 will be used.

19. Add 10 mg/ml tRNA to 20 μg/ml final, 1/10 vol of 3 M sodium acetate, pH 5.2 (*APPENDIX 2*), and 2.5 vol ice-cold 95% ethanol and place 15 min on dry ice. Microcentrifuge, wash and dry as in step 14, and resuspend pellet in 20 μl TE buffer.

    Ethanol precipitation also extracts the ethidium from the DNA.

20. Determine radioactivity in 1 μl using a fluor and scintillation counter and then calculate the recovery of double-stranded cDNA (see commentary). Proceed to library construction protocols in *UNIT 5.8*.

ALTERNATE PROTOCOL

LIGATION OF *Bst*XI SYNTHETIC ADAPTERS

Blunt-ended, double-stranded cDNA is ligated to phosphorylated *Bst*XI adapters and then purified as described in the basic protocol. Alternatively, *Eco*RI or *Eco*RI-*Not*I adapters may be used for cDNA to be cloned in vectors with the *Eco*RI site (Fig. 5.6.2). This protocol is simpler than that for linkers because the methylation and restriction digestion steps are unnecessary.

Additional Materials

*Bst*XI adapters (*UNIT 2.11*; Invitrogen), *Eco*RI adapters (New England Biolabs), *or* *Eco*RI-*Not*I adapters (Invitrogen)

1. Dissolve blunt-ended, double-stranded cDNA pellet in 23 μl water and add in the following order (30 μl final volume):

   3 μl 10× T4 DNA ligase buffer containing 5 mM ATP (1× and 0.5 mM ATP final)

   2 μl 1 μg/μl *Eco*RI, *Eco*RI-*Not*I, or phosphorylated *Bst*XI adapters (67 μg/ml final).

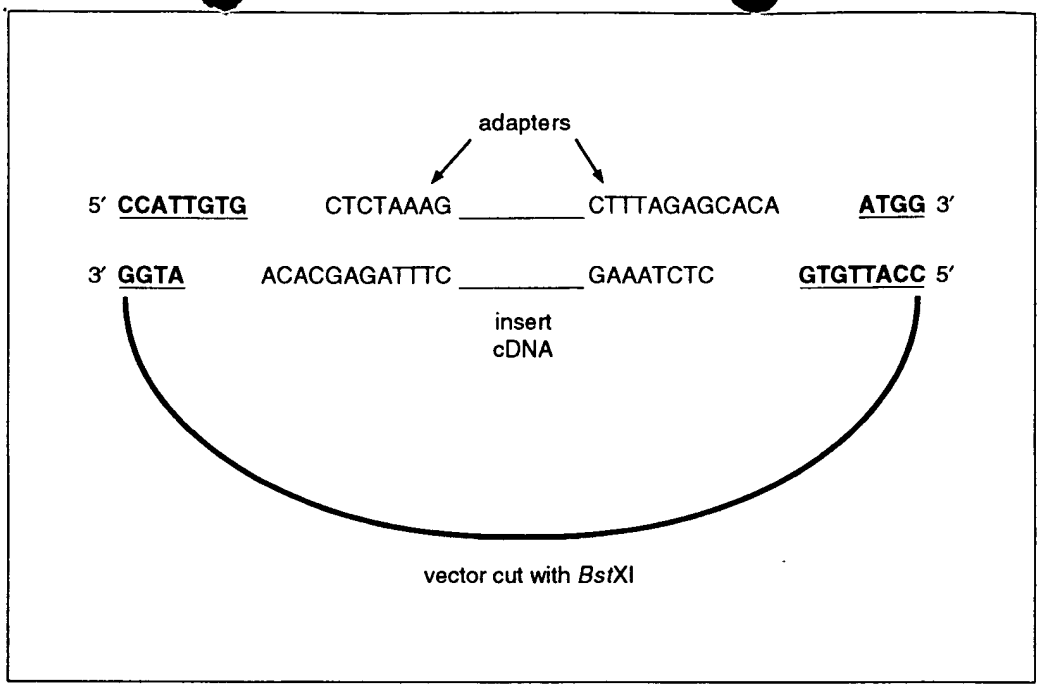


Figure 5.6.2 Noncomplementary adapter strategy. The insert and vector ends are compatible with each other but cannot self-ligate because the cohesive ends are not self-complementary. The *Bst*XI sites (underlined and bold) are not regenerated.

Mix gently by pipetting up and down with a pipettor. Add 2 µl (800 U) T4 DNA ligase to 27,000 U/ml final, mix as above, and incubate overnight at 4°C.

It is helpful to use ^{32}P-labeled cDNA to follow the DNA on the subsequent CL-4B column. If the cDNA is not ^{32}P-labeled, it may be labeled at this step by using [^{32}P]labeled adapters, prepared with [γ-^{32}P]ATP as in the T4 polynucleotide kinase exchange reaction (UNIT 3.10). Alternatively, if the adapters are not yet phosphorylated, they may be labeled with [γ-^{32}P]ATP as in the T4 polynucleotide kinase forward reaction (UNIT 3.10).

2. Add 100 µl TE buffer and remove excess adapters as in steps 10 to 15 of the basic protocol. If desired, size-select the cDNA as in steps 16 to 20 of the basic protocol. Resuspend purified cDNA pellet (obtained from either the CL-4B column or the gel) in 10 to 15 µl TE buffer. Proceed to library construction protocols in UNIT 5.8.

Because adapter dimers formed during the ligation reaction will clone into the vector very efficiently, removal of the excess adapters is essential.

PREPARATION OF A CL-4B COLUMN

SUPPORT PROTOCOL

The CL-4B column (Fig. 5.6.3) effectively removes linkers or adapters that would otherwise interfere in subsequent cloning steps; it also allows selection of cDNA ≥350 bp (see basic and alternate protocols). The column may be prepared while the cDNA is being digested with *Eco*RI, as described in steps 7 and 8 of the basic protocol.

Additional Materials

Preswollen Sepharose CL-4B (Pharmacia), 4°C
CL-4B column buffer
Silanized glass wool
Plastic tubing (new) with clamp

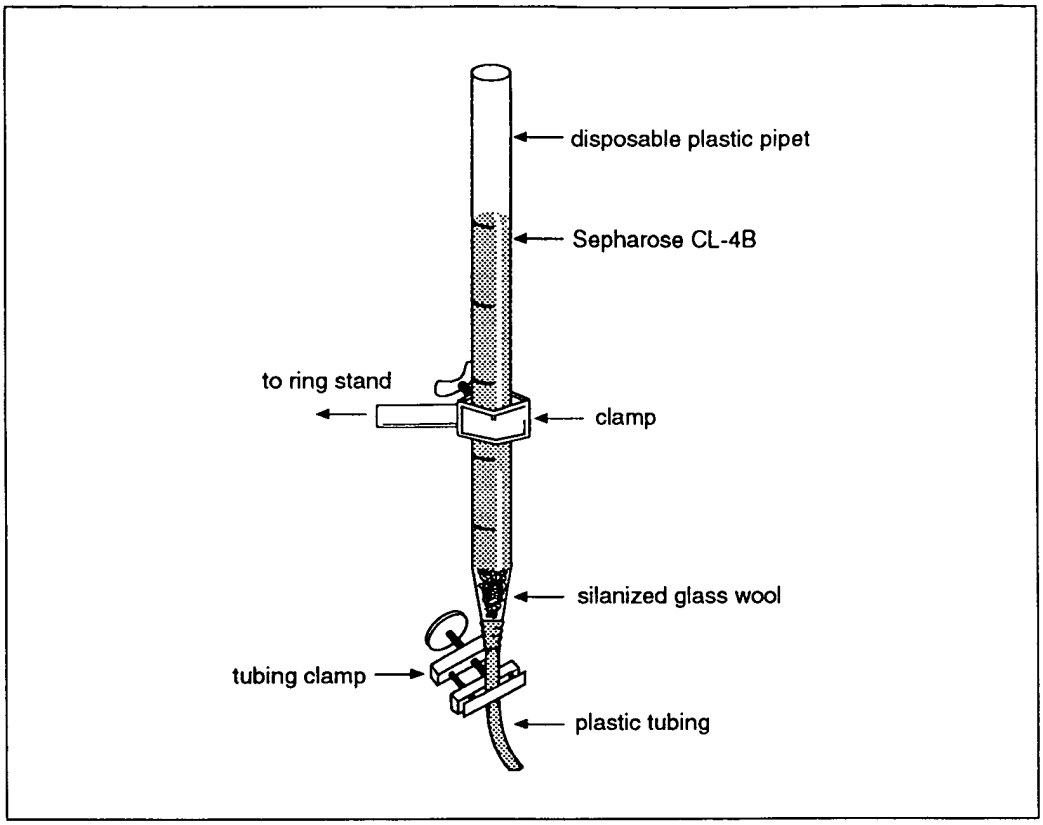


Figure 5.6.3 CL-4B column used for removal of *Eco*RI-digested linkers and selection of cDNA ≥350 bp.

1. Transfer 10 ml of preswollen Sepharose CL-4B to a 50-ml polypropylene tube and fill the tube with CL-4B column buffer. Mix by inverting several times and let the Sepharose CL-4B settle by gravity for 10 to 15 min. Aspirate the buffer above the settled gel, removing also the unsettled "fines."
2. Fill the tube two times with CL-4B column buffer—allow the Sepharose CL-4B to settle each time and remove the fines as in step 1.
3. Add 10 ml CL-4B column buffer and mix by inverting several times. Incubate the tube 10 min at 37°C, then proceed at room temperature.

   Outgassing may occur if the column is poured cold. The bubbles thus formed in the gel will interfere with the chromatography.
4. Break off the top of the 5-ml plastic pipet. Wearing gloves, use the 1-ml pipet to push a small piece (3- to 4-mm^3) of silanized glass wool down to the tip of the 5-ml pipet. Push a 3-cm length of plastic tubing firmly onto the tip of the 5-ml pipet. Clamp the tubing and attach the column to the ring stand as shown in Figure 5.6.3.
5. With a pipet, carefully fill the column with the gel slurry from step 3. After a few minutes, release the clamp on the tubing and allow the column to flow. Periodically add more slurry to the column as the level drops until the volume of packed gel in the column is at the 5-ml mark.
6. Allow the level of buffer in the column to drop until it is just above the level of the gel and clamp the tubing to stop the flow. The column is ready to be loaded (see basic or alternate protocols).

REAGENTS AND SOLUTIONS

Buffer stock solutions

Prepare ≥10 ml of each of the following stock solutions. Use autoclaved water and pass each solution through a sterile 0.45-μm filter. Store at room temperature unless otherwise indicated.

1 M Tris·Cl, pH 8.0 and pH 7.5
0.5 M EDTA, pH 8.0
3 M sodium acetate, pH 5.2
5 M NaCl (prepare 100 ml)
1 M $MgCl_2$
7.5 M ammonium acetate
20% *N*-lauroylsarcosine (Sarkosyl)
1 M DTT; store at –20°C in tightly capped tube
0.1 M ATP, pH 7.0 (prepare 1.0 ml); neutralize as described in *UNIT 3.4*; store at –20°C

From these stock solutions, prepare the following buffers, which should be checked for nuclease activity as described in *UNIT 5.5*. Enzyme buffers should be frozen in 200-μl aliquots at –80°C in screw-cap microcentrifuge tubes. Enzyme buffers prepared and stored as described will last for years.

Several of these solutions are routine and may be already available. Nonetheless, to help ensure success, it is best to prepare separate stocks for critical applications such as library preparation.

CL-4B column buffer, 500 ml

5 ml 1 M Tris·Cl, pH 8.0
60 ml 5 M NaCl
1 ml 0.5 M EDTA, pH 8.0
2.5 ml 20% Sarkosyl
431.5 ml H_2O
Filter sterilize and store at room temperature

50× S-adenosylmethionine (SAM)

1 mg SAM
1.0 ml 50× SAM dilution buffer (see below)

Prepare fresh just prior to use. Store dry SAM at –80°C for no longer than 2 months.

50× SAM dilution buffer, 7 ml

330 μl 3 M sodium acetate, pH 5.2
6.67 ml H_2O
Store in 1-ml aliquots

Silanized glass wool

Submerge the glass wool in a 1:100 dilution of a silanizing agent such as Prosil 28 (VWR) for 15 sec with shaking. Rinse the glass wool extensively with distilled H_2O. Autoclave the glass wool for 10 min and store at room temperature.

COMMENTARY

Background Information

The most common method currently employed to create compatible ends on cDNA prior to cloning is the attachment of synthetic linkers (basic protocol). This has for the most part superseded homopolymeric tailing (Maniatis et al., 1982), since linkering is relatively efficient, and its use eliminates the need for the sometimes tricky procedure of titrating the tailing reaction conditions. Other methods, such as the sequential ligation of two different linkers (Maniatis et al., 1982), are now rarely used. The linkered cDNA may be cloned into a plasmid or a phage vector. The use of synthetic adapters that ligate at high efficiency instead of linkers eliminates the methylation and restriction digestion steps (alternate protocol). The recently developed noncomplementary adapter strategy (see below) may ultimately supersede all the above methods as vectors with appropriate sites become available.

Size fractionation of cDNA is preferable to size fractionation of the initial mRNA primarily because DNA is considerably less susceptible to degradation than is RNA. In addition, the small fragments that are generated during the cDNA synthesis process (e.g., due to contaminating nucleases) are removed. These would otherwise be preferentially inserted and decrease the yield of long cDNA clones in the library. RNA enrichment procedures were undertaken in the past because cloning efficiencies then were not high enough to ensure the representation of very rare mRNAs in the library. Currently, the most common reason for using mRNA fractionation or enrichment is in preparation of subtracted libraries (*UNIT 5.8*; Hedrick et al., 1984).

The strategy of noncomplementary adapters (Fig. 5.6.2), developed by Brian Seed and coworkers (Seed, 1987), overcomes the loss of library complexity due to self-ligation of the cDNA inserts or the vector. This enables a high efficiency for the ligation of the vector to the insert since there are no competing reactions and the vector ends are similarly noncomplementary and do not self-ligate. The much higher yield of desirable ligation products ultimately results in a greater number of clones in the library. Furthermore, the occasional "scrambles" produced as a consequence of two unrelated cDNAs ligating together during the vector ligation step are eliminated; however, "scrambles" at the stage of adapter ligation remain possible, though infrequent.

Adapters are favored over linkers in general because the methylation and restriction digestion steps are bypassed, thus simplifying the procedure. adapters are required in the alternate protocol procedure to ensure that all of the cDNA will have the proper, noncomplementary "sticky ends." The use of *Eco*RI-*Not*I adapters for cloning with *Eco*RI compatible vectors introduces the rare *Not*I site at both ends of the insert, allowing the insert to be cut out of the vector as one fragment, which is not always possible with *Eco*RI linkers if there are internal *Eco*RI sites present.

The development of noncomplementary adapters has enabled the production of high-complexity cDNA libraries in multifunctional plasmid vectors such as CDM8 (Seed, 1987). These vectors permit library screening by functional expression in eukaryotic cells and production of single-stranded DNA for mutagenesis or subtraction, in addition to conventional hybridization methods. For some specialized applications and for selectable markers other than the *supF* present in CDM8, other vectors that employ the noncomplementary linker strategy (and use the same *Bst*XI sequence) include several available from Invitrogen with different antibiotic resistances and eukaryotic selectable markers. Other available vectors include AprM8, an ampicillin-resistant version of CDM8 (L.B. Klickstein, unpublished results), and retroviral vectors in the pBabe series (Morgenstern and Land, 1990).

Critical Parameters

In order to maximize the length and cloning efficiency of the cDNA, it is essential that contamination by endo- and exonucleases be avoided. EDTA, an inhibitor of most nucleases, should be present whenever feasible. Reagents, solutions, and enzymes should be of the highest quality obtainable.

The SAM reagent in the methylation step is very unstable, must be freshly dissolved prior to use, and should not be kept >2 months, even if stored dry at −80°C and never thawed. Stabilized SAM is supplied free of charge with methylases purchased from New England Biolabs.

Methylation conditions may be checked by methylating λ DNA under the conditions described in step 1 of the basic protocol, digesting methylated and unmethylated DNA with

*Eco*RI, and comparing the two samples by agarose gel electrophoresis. Methylated λ DNA should be protected from *Eco*RI digestion.

Digested linkers must be completely removed. Because of their small size, even as a small weight percentage of the total DNA, they will comprise a large mole fraction; thus, if not completely removed, most clones in the library will contain only linkers. Some investigators remove digested linkers by three sequential precipitations from 2 M ammonium acetate. However, column chromatography provides better resolution of the cDNA from the linkers. The CL-4B column also removes small double-stranded cDNA molecules. Purification of cDNA by gel electrophoresis after ligation of the linkers or adapters may be performed as an alternative to the CL-4B column, but the yield usually is not as high and the separation of linkers or adapters from cDNA is not as complete.

If *noncomplementary* adapters are employed, both 5′ ends should be phosphorylated for optimum results. If *complementary* adapters are used, only the blunt end should have a 5′ phosphate. The other end is phosphylated after ligation to the cDNA.

Because *Bst*XI produces a cohesive end with a sequence not specified by the restriction site sequence, the adapters and vector must correspond—e.g., the cohesive ends of the adapters used for the cDNA must be complementary to that of the vector. The vectors CDM8, AprM8, pCDNAI, pCDNAII, and the pBabe series all use the same sticky ends (Fig. 5.6.2).

Troubleshooting

Methylation and addition of linkers to double-stranded DNA are important to the successful construction of a cDNA library, but cannot be evaluated immediately. Unmethylated cDNA will clone efficiently and the problem will only be detected if the isolated inserts all end at an internal *Eco*RI site. This possibility may be minimized by prior evaluation of the quality of the SAM used in the methylation reaction (see reagents and solutions), or by use of the stabilized SAM that accompanies New England Biolab's *Eco*RI methylase.

Incomplete addition of linkers or adapters will be detected as a poor cloning efficiency in the next step of ligating the inserts to the vector. The problem may be any of the following: (1) cDNA was not blunted properly; (2) linkers or adapters were not kinased well or not annealed; (3) multiple linkers were not cut off with *Eco*RI; (4) ligation reaction did not work. However, the problem is usually improperly blunted cDNA or linkers that do not ligate well. Evaluate the linkered cDNA (after the CL-4B column) by ligating 5% of the sample with no vector and running the reaction on a 1% agarose minigel next to an equal amount of unligated cDNA. Properly linkered cDNA should significantly increase in size as determined by ethidium bromide staining or autoradiography of the dried gel. Poorly linkered cDNA may often be salvaged by repeating the blunting and linkering steps with fresh reagents.

Inadequate separation of linkers or adapters from cDNA will be detected as cloning efficiency that is too high and as clones with no inserts detectable by gel electrophoresis (white plaques with no inserts in λgt11 or plaques on C600hflA with λgt10 that have no insert). The remainder of this cDNA may be salvaged by repeating the CL-4B column chromatography.

Anticipated Results

Approximately 50% to 70% of the starting radioactivity present in the blunt-ended, double-stranded cDNA should be collected after the CL-4B column. This typically represents 1 to 3 μg of cDNA. In a human tonsil library prepared in λgt11 where the inserts were thoroughly evaluated, the mean insert size was 1.4 kb and actin clones represented 0.34% of all recombinants. Three-quarters of the actin cDNAs were nearly full length (actin mRNA = 2.1 kb).

Time Considerations

The methylation can be done on the same day as the blunting step from the cDNA synthesis protocol. The linker ligation is then set up overnight. *Eco*RI digestion and CL-4B chromatography are performed the following day, and the cDNA is either ligated to the vector overnight (*UNIT 5.8*) or stored as an ethanol precipitate overnight and further size selected by agarose gel electrophoresis the next day. At any ethanol precipitation in the procedure, the cDNA may be stored for several days as an ethanol precipitate.

The use of adapters rather than linkers eliminates the need for the methylation and linker digestion steps. The adapted cDNA is immediately loaded onto the CL-4B column after the adapter ligation and is used directly from the column after ethanol precipitation. The subsequent ligation of the adapted cDNA and the vector is performed overnight, requir-

ing the same amount of time with adapters as with linkers.

Literature Cited

Hedrick, S.M., Cohen, D.I., Nielsen, E.A., and Davis, M.M. 1984. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature (Lond.)* 308:149-153.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Morgenstern, J.P. and Land, H. 1990. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucl. Acids Res.* 18:3587-3596.

Seed, B. 1987. An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. *Nature (Lond.)* 329:840-842.

Key References

Neve, R.L., Hanis, P., Kosik, K.S., Kurnit, D.M., and Donlon, T.A. 1986. Identification of the gene for the human microtubule-associated protein tau and chromosomal localization of the genes for tau and microtubule-associated protein 2. *Brain Res.* 387:271-80.

Seed, B. 1987. See above.

Contributed by Lloyd B. Klickstein
Brigham and Women's Hospital
Boston, Massachusetts

Rachael L. Neve
McLean Hospital
Belmont, Massachusetts